OIPE

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/040,647

DATE: 05/07/2002
TIME: 11:16:47

Input Set : N:\Crf3\RULE60\10040647.raw
Output Set: N:\CRF3\05072002\J040647.raw

## SEQUENCE LISTING

```
 3 (1) GENERAL INFORMATION:
 5     (i) APPLICANT: (US only) ANTALIS Toni Marie and HOOPER John David
 7    (ii) TITLE OF INVENTION: NOVEL MOLECULES
 9   (iii) NUMBER OF SEQUENCES: 30
11    (iv) CORRESPONDENCE ADDRESS:
12        (A) ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
13        (B) STREET: 400 GARDEN CITY PLAZA
14        (C) CITY: GARDEN CITY
15        (D) STATE: NEW YORK
16        (E) COUNTRY: USA
17        (F) ZIP: 11530
19     (v) COMPUTER READABLE FORM:
20        (A) MEDIUM TYPE: Floppy disk
21        (B) COMPUTER: IBM PC compatible
22        (C) OPERATING SYSTEM: PC-DOS/MS-DOS
23        (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
25    (vi) CURRENT APPLICATION DATA:
C-->26        (A) APPLICATION NUMBER: US/10/040,647
C-->27        (B) FILING DATE: 07-Jan-2002
28        (C) CLASSIFICATION:
42   (vii) PRIOR APPLICATION DATA:
31        (A) APPLICATION NUMBER: 09/023,942
32        (B) FILING DATE:
36        (A) APPLICATION NUMBER: PO5101/97
37        (B) FILING DATE: 13-FEB-1997
40        (A) APPLICATION NUMBER: PP0422/97
41        (B) FILING DATE: 18-NOV-1997
43        (A) APPLICATION NUMBER: International PCT Application
44        (B) FILING DATE: 13-FEB-1998
46  (viii) ATTORNEY/AGENT INFORMATION:
47        (A) NAME: DIGIGLIO, FRANK S
48        (B) REGISTRATION NUMBER: 31,346
49        (C) REFERENCE/DOCKET NUMBER: 11168
51    (ix) TELECOMMUNICATION INFORMATION:
52        (A) TELEPHONE: (516) 742 4343
53        (B) TELEFAX: (516) 742 4366
54        (C) TELEX: 230 901 SANS UR
58 (2) INFORMATION FOR SEQ ID NO: 1:
60     (i) SEQUENCE CHARACTERISTICS:
61        (A) LENGTH: 32 base pairs
62        (B) TYPE: nucleic acid
63        (C) STRANDEDNESS: single
```

ENTERED

RAW SEQUENCE LISTING          DATE: 05/07/2002
PATENT APPLICATION:  US/10/040,647     TIME: 11:16:47

Input Set : N:\Crf3\RULE60\10040647.raw
Output Set: N:\CRF3\05072002\J040647.raw

```
      64              (D) TOPOLOGY: linear
W--> 66        (ii) MOLECULE TYPE: DNA
      68        (ix) FEATURE:
      69              (A) NAME/KEY: modified base
      70              (B) LOCATION: 15
      71              (D) OTHER INFORMATION: N equals Inosine
      73        (ix) FEATURE:
      74              (A) NAME/KEY: modified base
      75              (B) LOCATION: 18
      76              (D) OTHER INFORMATION: N equals Inosine
      78        (ix) FEATURE:
      79              (A) NAME/KEY: modified base
      80              (B) LOCATION: 21
      81              (D) OTHER INFORMATION: N equals Inosine
      84        (ix) FEATURE:
      85              (A) NAME/KEY: modified base
      86              (B) LOCATION: 24
      87              (D) OTHER INFORMATION: N equals Inosine
      90        (ix) FEATURE:
      91              (A) NAME/KEY: modified base
      92              (B) LOCATION: 27
      93              (D) OTHER INFORMATION: N equals Inosine
      95     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
      97 ACAGAATTCT GGGTNGTNAC NGCNGCNCAY TG                          32
     100 (2) INFORMATION FOR SEQ ID NO: 2:
     102        (i) SEQUENCE CHARACTERISTICS:
     103              (A) LENGTH: 29 base pairs
     104              (B) TYPE: nucleic acid
     105              (C) STRANDEDNESS: single
     106              (D) TOPOLOGY: linear
W--> 108        (ii) MOLECULE TYPE: DNA
     110        (ix) FEATURE:
     111              (A) NAME/KEY: modified base
     112              (B) LOCATION: 12
     113              (D) OTHER INFORMATION: N equals Inosine
     115        (ix) FEATURE:
     116              (A) NAME/KEY: modified base
     117              (B) LOCATION: 15
     118              (D) OTHER INFORMATION: N equals Inosine
     120        (ix) FEATURE:
     121              (A) NAME/KEY: modified base
     122              (B) LOCATION: 18
     123              (D) OTHER INFORMATION: N equals Inosine
     125        (ix) FEATURE:
     126              (A) NAME/KEY: modified base
     127              (B) LOCATION: 21
     128              (D) OTHER INFORMATION: N equals Inosine
     130        (ix) FEATURE:
     131              (A) NAME/KEY: modified base
```

RAW SEQUENCE LISTING                    DATE: 05/07/2002
PATENT APPLICATION:  US/10/040,647      TIME: 11:16:47

Input Set : N:\Crf3\RULE60\10040647.raw
Output Set: N:\CRF3\05072002\J040647.raw

```
132            (B) LOCATION: 27
133            (D) OTHER INFORMATION: N equals Inosine
135      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
137 ACAGAATTCA RNGGNCCNCC NSWRTCNCC                                    29
140 (2) INFORMATION FOR SEQ ID NO: 3:
142      (i) SEQUENCE CHARACTERISTICS:
143            (A) LENGTH: 1094 base pairs
144            (B) TYPE: nucleic acid
145            (C) STRANDEDNESS: single
146            (D) TOPOLOGY: linear
W--> 148     (ii) MOLECULE TYPE: DNA
150      (ix) FEATURE:
151            (A) NAME/KEY: CDS
152            (B) LOCATION: 17..955
154      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
156 CGCGGGAGAG GAGGCC ATG GGC GCG CGC GGG GCG CTG CTG CTG GCG CTG     49
157                    Met Gly Ala Arg Gly Ala Leu Leu Leu Ala Leu
158                     1               5                  10
160 CTG CTG GCT CGG GCT GGA CTC AGG AAG CCG GAG TCG CAG GAG GCG GCG   97
161 Leu Leu Ala Arg Ala Gly Leu Arg Lys Pro Glu Ser Gln Glu Ala Ala
162           15                  20                  25
164 CCG TTA TCA GGA CCA TGC GGC CGA CGG GTC ATC ACG TCG CGC ATC GTG  145
165 Pro Leu Ser Gly Pro Cys Gly Arg Arg Val Ile Thr Ser Arg Ile Val
166           30                  35                  40
168 GGT GGA GAG GAC GCC GAA CTC GGG CGT TGG CCG TGG CAG GGG AGC CTG  193
169 Gly Gly Glu Asp Ala Glu Leu Gly Arg Trp Pro Trp Gln Gly Ser Leu
170           45                  50                  55
172 CGC CTG TGG GAT TCC CAC GTA TGC GGA GTG AGC CTG CTC AGC CAC CGC  241
173 Arg Leu Trp Asp Ser His Val Cys Gly Val Ser Leu Leu Ser His Arg
174 60                  65                  70                  75
176 TGG GCA CTC ACG GCG GCG CAC TGC TTT GAA ACT GAC CTT AGT GAT CCC  289
177 Trp Ala Leu Thr Ala Ala His Cys Phe Glu Thr Asp Leu Ser Asp Pro
178           80                  85                  90
180 TCC GGG TGG ATG GTC CAG TTT GGC CAG CTG ACT TCC ATG CCA TCC TTC  337
181 Ser Gly Trp Met Val Gln Phe Gly Gln Leu Thr Ser Met Pro Ser Phe
182           95                  100                 105
184 TGG AGC CTG CAG GCC TAC TAC ACC CGT TAC TTC GTA TCG AAT ATC TAT  385
185 Trp Ser Leu Gln Ala Tyr Tyr Thr Arg Tyr Phe Val Ser Asn Ile Tyr
186           110                 115                 120
187 CTG AGC CCT CGC TAC CTG GGG AAT TCA CCC TAT GAC ATT GCC TTG GTG  433
188 Leu Ser Pro Arg Tyr Leu Gly Asn Ser Pro Tyr Asp Ile Ala Leu Val
189           125                 130                 135
191 AAG CTG TCT GCA CCT GTC ACC TAC ACT AAA CAC ATC CAG CCC ATC TGT  481
192 Lys Leu Ser Ala Pro Val Thr Tyr Thr Lys His Ile Gln Pro Ile Cys
193 140                 145                 150                 155
195 CTC CAG GCC TCC ACA TTT GAG TTT GAG AAC CGG ACA GAC TGC TGG GTG  529
196 Leu Gln Ala Ser Thr Phe Glu Phe Glu Asn Arg Thr Asp Cys Trp Val
197           160                 165                 170
199 ACT GGC TGG GGG TAC ATC AAA GAG GAT GAG GCA CTG CCA TCT CCC CAC  577
```

RAW SEQUENCE LISTING                          DATE: 05/07/2002
PATENT APPLICATION:  US/10/040,647           TIME: 11:16:47

Input Set : N:\Crf3\RULE60\10040647.raw
Output Set: N:\CRF3\05072002\J040647.raw

```
200 Thr Gly Trp Gly Tyr Ile Lys Glu Asp Glu Ala Leu Pro Ser Pro His
201         175                 180             185
203 ACC CTC CAG GAA GTT CAG GTC GCC ATC ATA AAC AAC TCT ATG TGC AAC        625
204 Thr Leu Gln Glu Val Gln Val Ala Ile Ile Asn Asn Ser Met Cys Asn
205         190                 195             200
207 CAC CTC TTC CTC AAG TAC AGT TTC CGC AAG GAC ATC TTT GGA GAC ATG        673
208 His Leu Phe Leu Lys Tyr Ser Phe Arg Lys Asp Ile Phe Gly Asp Met
209         205                 210             215
211 GTT TGT GCT GGC AAT GCC CAA GGC GGG AAG GAT GCC TGC TTC GGT GAC        721
212 Val Cys Ala Gly Asn Ala Gln Gly Gly Lys Asp Ala Cys Phe Gly Asp
213 220                 225             230                 235
215 TCA GGT GGA CCC TTG GCC TGT AAC AAG GAT GGA CTG TGG TAT CAG ATT        769
216 Ser Gly Gly Pro Leu Ala Cys Asn Lys Asp Gly Leu Trp Tyr Gln Ile
217             240                 245                 250
219 GGA GTC GTG AGC TGG GGA GTG GGC TGT GGT CGG CCC AAT CGG CCC GGT        817
220 Gly Val Val Ser Trp Gly Val Gly Cys Gly Arg Pro Asn Arg Pro Gly
221             255                 260                 265
223 GTC TAC ACC AAT ATC AGC CAC CAC TTT GAG TGG ATC CAG AAG CTG ATG        865
224 Val Tyr Thr Asn Ile Ser His His Phe Glu Trp Ile Gln Lys Leu Met
225         270                 275                 280
226 GCC CAG AGT GGC ATG TCC CAG CCA GAC CCC TCC TGG CCG CTA CTC TTT        913
227 Ala Gln Ser Gly Met Ser Gln Pro Asp Pro Ser Trp Pro Leu Leu Phe
228         285                 290                 295
230 TTC CCT CTT CTC TGG GCT CTC CCA CTC CTG GGG CCG GTC TGA            955
231 Phe Pro Leu Leu Trp Ala Leu Pro Leu Leu Gly Pro Val  *
232 300                 305                 310
234 GCCTACCTGA GCCCATGCAG CCTGGGGCCA CTGCCAAGTC AGGCCCTGGT TCTCTTCTGT     1015
236 CTTGTTTGGT AATAAACACA TTCCAGTTGA TGCCTTGCAG GGCATTTTTC AAAAAAAAAA     1075
238 AAAAAAAAAA AAAAAAAA                                                   1094
241 (2) INFORMATION FOR SEQ ID NO: 4:
243        (i) SEQUENCE CHARACTERISTICS:
244            (A) LENGTH: 312 amino acids
245            (B) TYPE: amino acid
246            (D) TOPOLOGY: linear
248        (ii) MOLECULE TYPE: protein
250        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
252 Met Gly Ala Arg Gly Ala Leu Leu Leu Ala Leu Leu Leu Ala Arg Ala
253   1           5                 10                  15
255 Gly Leu Arg Lys Pro Glu Ser Gln Glu Ala Ala Pro Leu Ser Gly Pro
256             20                 25                 30
258 Cys Gly Arg Arg Val Ile Thr Ser Arg Ile Val Gly Gly Glu Asp Ala
259         35                 40                 45
261 Glu Leu Gly Arg Trp Pro Trp Gln Gly Ser Leu Arg Leu Trp Asp Ser
262     50                 55                 60
264 His Val Cys Gly Val Ser Leu Leu Ser His Arg Trp Ala Leu Thr Ala
265 65                 70                 75                 80
267 Ala His Cys Phe Glu Thr Asp Leu Ser Asp Pro Ser Gly Trp Met Val
268             85                 90                 95
270 Gln Phe Gly Gln Leu Thr Ser Met Pro Ser Phe Trp Ser Leu Gln Ala
```

RAW SEQUENCE LISTING
PATENT APPLICATION:  US/10/040,647

DATE: 05/07/2002
TIME: 11:16:47

Input Set : N:\Crf3\RULE60\10040647.raw
Output Set: N:\CRF3\05072002\J040647.raw

```
271              100               105               110
273 Tyr Tyr Thr Arg Tyr Phe Val Ser Asn Ile Tyr Leu Ser Pro Arg Tyr
274         115               120               125
276 Leu Gly Asn Ser Pro Tyr Asp Ile Ala Leu Val Lys Leu Ser Ala Pro
277     130               135               140
279 Val Thr Tyr Thr Lys His Ile Gln Pro Ile Cys Leu Gln Ala Ser Thr
280 145               150               155               160
282 Phe Glu Phe Glu Asn Arg Thr Asp Cys Trp Val Thr Gly Trp Gly Tyr
283             165               170               175
285 Ile Lys Glu Asp Glu Ala Leu Pro Ser Pro His Thr Leu Gln Glu Val
286         180               185               190
288 Gln Val Ala Ile Ile Asn Asn Ser Met Cys Asn His Leu Phe Leu Lys
289         195               200               205
291 Tyr Ser Phe Arg Lys Asp Ile Phe Gly Asp Met Val Cys Ala Gly Asn
292     210               215               220
294 Ala Gln Gly Gly Lys Asp Ala Cys Phe Gly Asp Ser Gly Gly Pro Leu
295 225               230               235               240
297 Ala Cys Asn Lys Asp Gly Leu Trp Tyr Gln Ile Gly Val Val Ser Trp
298             245               250               255
300 Gly Val Gly Cys Gly Arg Pro Asn Arg Pro Gly Val Tyr Thr Asn Ile
301             260               265               270
303 Ser His His Phe Glu Trp Ile Gln Lys Leu Met Ala Gln Ser Gly Met
304         275               280               285
306 Ser Gln Pro Asp Pro Ser Trp Pro Leu Leu Phe Phe Pro Leu Leu Trp
307     290               295               300
309 Ala Leu Pro Leu Leu Gly Pro Val
310 305               310
```

313 (2) INFORMATION FOR SEQ ID NO: 5:
315        (i) SEQUENCE CHARACTERISTICS:
316            (A) LENGTH: 1100 base pairs
317            (B) TYPE: nucleic acid
318            (C) STRANDEDNESS: single
319            (D) TOPOLOGY: linear
W--> 321       (ii) MOLECULE TYPE: DNA
323        (ix) FEATURE:
324            (A) NAME/KEY: CDS
325            (B) LOCATION: 17..961
327        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
329 CGCGGGAGAG GAGGCC ATG GGC GCG CGC GGG GCG CTG CTG CTG GCG CTG            49
330                    Met Gly Ala Arg Gly Ala Leu Leu Leu Ala Leu
331                     1               5               10
333 CTG CTG GCT CGG GCT GGA CTC AGG AAG CCG GAG TCG CAG GAG GCG GCG        97
334 Leu Leu Ala Arg Ala Gly Leu Arg Lys Pro Glu Ser Gln Glu Ala Ala
335         15               20               25
337 CCG TTA TCA GGA CCA TGC GGC CGA CGG GTC ATC ACG TCG CGC ATC GTG      145
338 Pro Leu Ser Gly Pro Cys Gly Arg Arg Val Ile Thr Ser Arg Ile Val
339         30               35               40
341 GGT GGA GAG GAC GCC GAA CTC GGG CGT TGG CCG TGG CAG GGG AGC CTG      193
342 Gly Gly Glu Asp Ala Glu Leu Gly Arg Trp Pro Trp Gln Gly Ser Leu
```

**VERIFICATION SUMMARY**
PATENT APPLICATION:  **US/10/040,647**

DATE: 05/07/2002
TIME: 11:16:48

Input Set : **N:\Crf3\RULE60\10040647.raw**
Output Set: **N:\CRF3\05072002\J040647.raw**

```
L:26 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:27 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:66 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=1
L:108 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=2
L:148 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=3
L:321 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=5
L:406 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:495 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=7
L:641 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=9
L:928 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=11
L:944 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=12
L:959 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=13
L:973 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=14
L:1030 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=18
L:1043 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=19
L:1057 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=20
L:1071 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=21
L:1085 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=22
L:1099 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=23
L:1128 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=25
L:1287 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=27
L:1430 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=28
L:1483 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=29
L:1527 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=30
```